DECLARATION OF KONSTANTINOS PSOUNIS ISO GOOGLE LLC'S RESPONSE TO THE COURT'S 10/27/22 ORDER TO SHOW CAUSE (DKT. 784)

Redacted Version of Document Sought to be Sealed

HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA, OAKLAND DIVISION

CHASOM BROWN, et al., on behalf of themselves and all others similarly situated, Plaintiffs, vs. GOOGLE LLC, Defendant.		Case No. 4:20-cv-03664-YGR-SVK

DECLARATION OF KONSTANTINOS PSOUNIS, PHD

1. I have been retained by Google to evaluate certain logs disclosed by Google in this litigation to assess (i) whether or not a [redacted] log joins authenticated data with unauthenticated data; and (ii) whether or not any of the additional logs that Google disclosed in this litigation on June 14, 2022 would change any of the opinions in the expert report I submitted in this matter on June 7, 2022.

2. As explained further below, I have reached the following opinions: (i) [redacted] does not join authenticated data with unauthenticated data; and (ii) the existence of the additional logs Google disclosed on June 14, 2022 would not change any of the opinions stated in my June 7, 2022 Expert Report.

3. In reaching these opinions, I reviewed the following information: Google employee declarations and exhibits thereto that Google is submitting concurrently with this declaration, source code related to the [redacted], Plaintiffs' August 4, 2022 (Dkt. 655-1) and August 25, 2022 (Dkt. 707-1) briefs related to their request for supplemental sanctions, Exhibit B attached to this declaration, and publicly-available technical literature

regarding certain code functions. Google provided me with all information I asked for to enable me to render the opinions in this declaration.

BACKGROUND AND QUALIFICATIONS

4. I am a Professor and Associate Chair of Electrical and Computer Engineering and Professor of Computer Science at the University of Southern California. I joined the University of Southern California in 2003, after completing my PhD at Stanford University as a Stanford Graduate Fellow. Attached hereto as Exhibit A is a true and correct copy of my curriculum vitae.

5. My professional career has spanned more than 20 years. As set forth in Exhibit A, I have extensive experience in the field of networked distributed systems, including the Internet and the world wide web, content-delivery networks, data centers and cloud computing, and wireless mobile networking systems. Throughout my career, I have analyzed, designed, and developed efficient, privacy-preserving networked distributed systems for the Internet and the Web, content-delivery networks, data centers and cloud systems, and wireless mobile networking systems. As such, I have acquired extensive expertise in the analysis and development of those systems and the source code on which they rely.

6. I have published more than 100 technical papers in the field of networked distributed systems, which have been cited tens of thousands of times. I have also been awarded numerous grants and significant funding from the government and industry leaders to advance these fields. As a result, I have been named an Institute of Electrical and Electronics Engineers (IEEE) Fellow, the highest grade of membership, and a Distinguished Member of the Association of Computing Machinery (ACM) for my contributions to the theory and practice of networked, distributed systems.

OPINIONS

No Joining of Log Entries

7. I have reviewed Plaintiffs' Motion for Supplemental Sanctions (Dkt. 655-1) and Reply In Support Thereof (Dkt. 707-1). I understand that Plaintiffs have argued that the "joined logs" identified in the June 14, 2022 Declaration of Martin Šrámek are "relevant to key issues in this case, including . . . joining [private browsing data] with authenticated data," and if these logs had been disclosed earlier, Plaintiffs could have sought "additional discovery regarding . . . how Google joins private browsing data with authenticated data." Dkt. 655-1 at 2. According to Plaintiffs, "these additional logs contain highly relevant data . . . including . . . private browsing data that Google joined with authenticated data." *Id.* at 2–3. I also understand that Plaintiffs have argued that these logs show that "data flagged with an Incognito-detection bit in such logs *can* be joined with users' 'authenticated' data to identify them." Dkt. 707-1 at 3.

8. I have also reviewed the declarations of Vasily Panferov and Eugene Lee submitted concurrently with this declaration.

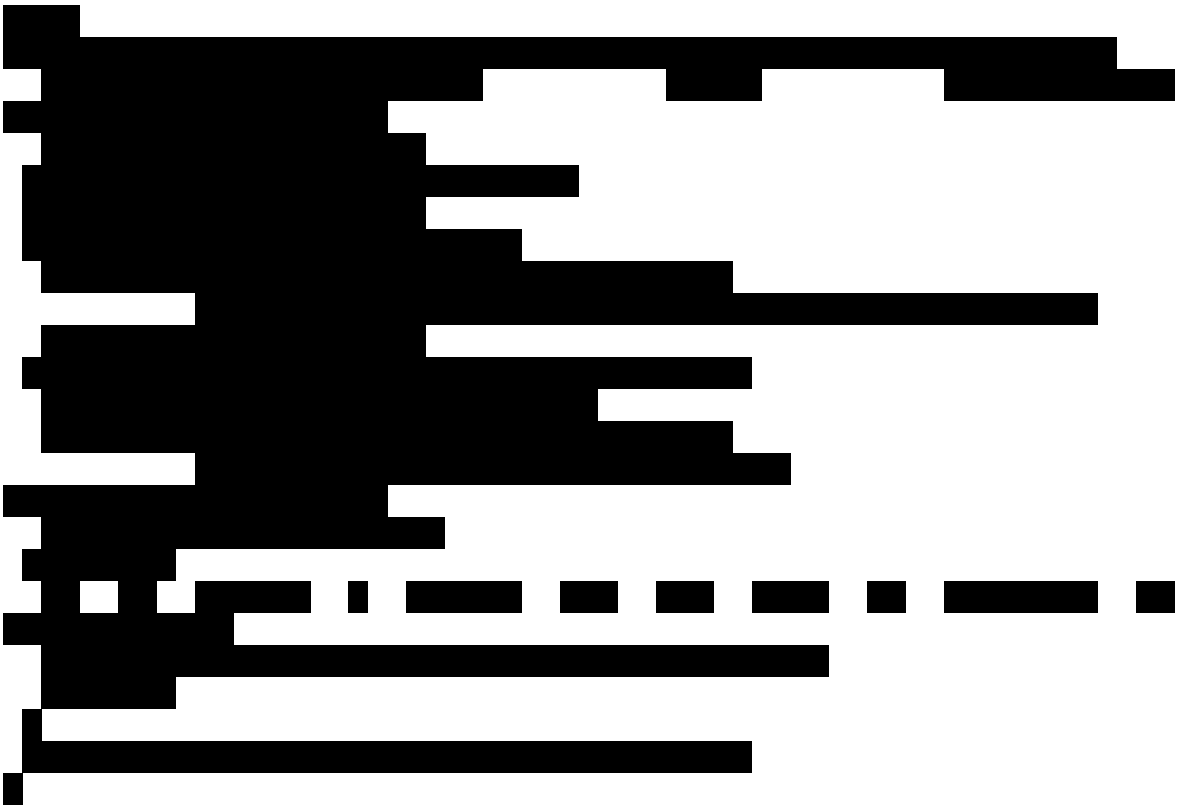
9. Based on my experience and training, "joining" of log data refers to associating separate log records with a shared key, such as an identifier. For the logs in question, authenticated and unauthenticated data would be considered "joined" if a log shows that a shared key (or any common data point) was used to associate or combine unauthenticated private browsing data at issue with an individual's Google account.

10. I have examined a source code file called [redacted], which contains the instructions for sorting input logs data in the [redacted] log. I requested this file from counsel after it was identified for me by Mr. Panferov. I have also discussed this source code with Mr. Panferov. Based on my review, it is my opinion that unauthenticated data is not joined with authenticated data in [redacted].

HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY

11. Below, I discuss in detail the key steps in adding inputs to generate this log, which support my opinion.

12. The first step is to call the function [redacted], which is used to extract from an authenticated log entry the corresponding GAIA ID, or, if the log entry is unauthenticated, the corresponding Zwieback ID. The full body of the function is reproduced here:



13. An "if" statement controls whether or not a section of a program is performed, based on evaluation of a condition contained in the parentheses following the "if" command. If the condition is evaluated to be true, the section of the program contained in brackets following the parentheses will be performed. If the condition is evaluated to be false, that section of the program will not be performed. In the code snippet reproduced in paragraph 12, the first "if" statement (which I have reproduced in red font below) prescribes a logic where (i) if a GAIA ID is contained in a log entry, then (ii) the GAIA ID is stored in a variable called [redacted], and a variable called [redacted], which indicates whether this is an authenticated log entry, is set to true:

[redacted]

Because this is an "if" statement conditioned on presence of a GAIA ID, it will only run when a GAIA ID is present in a given log entry.

14. If the condition in a preceding "if" statement is false, an "else if" statement can be used to specify a new condition to control whether or not another section of the program is performed. If the condition for the original "if" statement is evaluated as true (and thus the corresponding portion of the program is performed), any "else if" statement that follows that portion of the program will not be performed. In other words, the "else if" condition is only evaluated when the preceding "if" condition is false. In the code snippet reproduced in paragraph 12, the "else if" statement (which I have reproduced in blue font below) prescribes a logic where (i) if (and only if) a Zwieback ID (and not a GAIA ID) is contained in a log entry, then (ii) the Zwieback ID is stored in the [redacted] variable, and the [redacted] variable, which indicates whether this is an authenticated log entry, is set to false.

[redacted]

Because this is an "else if" statement, it will only run when the condition described in paragraph 13 is not satisfied (*i.e.*, if there is ***no GAIA ID*** present in a given log entry), but a Zwieback ID is present in that log entry.

15. The second step in generating the [redacted] log is to call the function [redacted] which is used to (i) generate a new key which combines

the [redacted] with the [redacted] for the log entry, and then (ii) assign the value of this key to the variable [redacted]. The code snippet below calls the function:

[redacted]

and the code snippet below assigns the value of the new key to the variable [redacted]:

[redacted]

A [redacted]. Because, for all practical purposes, each log entry has a unique [redacted] this second step guarantees that it is practically impossible for two log entries to have the same [redacted]. For example, even if two log entries have the same GAIA ID, they will have a different [redacted] and because the value of the [redacted] is based on the combination of the [redacted] (in this example, the GAIA ID) and the [redacted], these two log entries will have different [redacted] values. Similarly, even if two log entries have the same Zwieback ID, they will have a different [redacted] and, as a result, these two log entries will have different [redacted] values. Entries keyed to a Zwieback ID will never share a [redacted] with entries keyed to a GAIA ID.

16. The third and last step in generating the [redacted] log is to call the function [redacted] which is used to sort the log entries based on the [redacted] value, via the code snippet below:

[redacted]

The function [redacted] belongs to the well known [redacted] framework.[1] Mr. Panferov identified a document which describes this function. I asked for that document and it was provided to me by counsel; it is attached hereto as Exhibit B. Once sorted by the [redacted], the corresponding log entries are written to the log [redacted] one by one, as separate entries.

17. Based on my review of this code and Exhibit B, it is evident that there is no joining of any log entries during the operation described above. The [redacted] log lists log entries from three authenticated and three unauthenticated logs one by one, as separate entries, based on the aforementioned sorting. No two log entries from the aforementioned logs are "joined" because records from the input logs are not combined together into a single record (*see* Figure 1, *infra*). In the extremely unlikely event that two log entries with the same GAIA ID also have the same [redacted] then these two authenticated entries could be joined. Similarly, in the extremely unlikely event that two log entries with the same Zwieback ID also have the same [redacted] then these two unauthenticated entries could be joined. However, there is no circumstance in which an authenticated log entry could be joined with an unauthenticated log entry, because the GAIA ID of the former will never match the Zwieback ID of the latter.

18. In summary, (i) it is extremely unlikely that any authenticated log entries will be joined together, (ii) it is extremely unlikely that any unauthenticated log entries will be joined together, and (iii) it is impossible for authenticated and unauthenticated log entries to be joined together. The code does not allow joining of authenticated data with unauthenticated data (or vice versa). My conclusion is based on the logic in the code itself, which is discussed above.

[1] *See* Jeffrey Dean and Sanjay Ghemawat, "[redacted]: simplified data processing on large clusters," OSDI '04: Sixth Symposium on Operating System Design and Implementation at 137-150 (2004).

19. An illustration of how this works in practice for [redacted] is below:

[redacted]

20. My conclusions in paragraphs 17 and 18 above are consistent with the opinions stated in my June 7, 2022 Expert Report submitted in this case regarding Google's policy and technical restrictions that prohibit joining of authenticated and unauthenticated data. *See, e.g.*, Dkt. 659-10 §§ III.A, III.C., III.F, III.G, IV.C. The declarations of Mssrs. Panferov and Lee, along with Exhibit B attached hereto, are also consistent with these opinions. As explained above, the source code that defines the joining logic for [redacted] maintains a separation between records containing authenticated data and records containing unauthenticated data.

21. Certainly nothing in these logs or the source code would permit joining a given user's authenticated browsing activity with his or her unauthenticated browsing activity.

Previously-Proffered Opinions Regarding Class Member Identification and (Lack of) Joining of Authenticated and Unauthenticated Data

22. I have also reviewed the May 31, 2022 ("May 31 Šrámek Declaration") and June 14, 2022 Declarations ("June 14 Šrámek Declaration") of Martin Šrámek served by Google in this litigation, as well as the six Declarations from Google employees about each of the [redacted] log categories spanning [redacted] additional logs identified in the June 14 Šrámek Declaration.[2]

23. The [redacted] additional logs containing the maybe_chrome_incognito bit and one additional log containing the is_chrome_non_incognito_mode bit identified in the June 14 Šrámek Declaration do not change any of the opinions stated in my June 7, 2022 Expert Report ("Report") submitted in this case. In particular, my Report includes several opinions regarding Plaintiffs' proposed method for identifying putative class members. *See* Report §§ III.A, C, F-H, J, and IV.C. For the reasons below, the existence of the additional [redacted] logs does not change any of the opinions I have offered in this case.

24. **Report Section III.A (Opinion 1): "Mr. Hochman's Opinion That Users Can Readily Be Identified From The Data At Issue (# 18) Is Incorrect."** The additional [redacted] logs do not change this opinion because the data-at-issue is orphaned and unidentified: (a) any unauthenticated data in these logs is still keyed to an unauthenticated identifier (or no identifier) for a signed-out user that is unique to the private browsing session; (b) these logs do not change the operation of the cookie jar and server-side processes for users in private browsing modes described in paragraphs 37 through 58 of my Report; and (c) in addition to the policy and technical restrictions described in Sections III.A, III.C., III.F, III.G, and IV.C of my report, the source code discussed

[2] The 46 additional logs include the 44 logs identified in Exhibit A to the June 14 Šrámek Declaration, one Oolong log identified in paragraph 12 of the June 14 Šrámek Declaration, and ads:tmp-TestUnmatchedSearchAdQueryState. *See* Maki Decl. n. 1.

above provides another example of technical solutions Google has implemented to prevent the joining of authenticated and unauthenticated data.

25. **Report Section III.C (Opinion 3): "Mr. Hochman's Opinions On 'Private Browsing Profiles,' Server-Side Processes, And Data Joinability (# 10, 18, 19, 20) Are Inaccurate."** The additional [redacted] logs also do not change this opinion because (i) they do not show that Google maintains "cradle-to-grave" profiles of users that join signed-out private browsing mode activity with a Google account; and (ii) they do not change Google's server-side processes designed to prevent the joining of authenticated and unauthenticated data. Moreover, as discussed above, the lone "joined log" included in these [redacted] logs that contains authenticated and unauthenticated records does not actually "join" any such records, and the code that I have analyzed provides another example of technical restrictions that Google has implemented to prevent such joining.

26. **Report Section III.F (Opinion 6) "Mr. Hochman's Assertions On Fingerprinting Are Misleading And Unfounded."** The additional [redacted] logs also do not change this opinion because they do not show that Google engages in fingerprinting or undermine any of my opinions regarding the technical and policy constraints that Google implements to prevent the use of fingerprinting to re-identify users. Additionally, the source code analyzed above provides an additional example of technical restrictions that Google has implemented in line with its policies prohibiting fingerprinting and/or joining of authenticated and unauthenticated data.

27. **Report Section III.G.1 (Opinion 7): "Plaintiffs' Proposed "IP + UA Fingerprinting Method [for identifying members of Class I] Will Not Work."** As explained further below, the additional [redacted] logs do not change this opinion because the same issues described in my Report will also affect any attempt to identify class members by applying the same fingerprinting methodology to these logs. *Id.* Specifically, the proposed IP + UA fingerprinting method would still rely on combinations of IPv4/IPv6 addresses and user agents that are "not

sufficiently unique to identify class members because there are many situations where more than one user will have an identical IP address and user agent" if it were applied to these [redacted] additional logs, and these [redacted] additional logs do not resolve this fundamental flaw because it is inherent to the proposed use of a combination of an IP address and user agent as a join key to identify class members. *Id.* ¶ 111.

28. **Report Section III.H (Opinion 8): "Mr. Hochman's Opinion That The 'maybe_chrome_incognito' Bit Reliably Detects Incognito Traffic (# 23) Is Incorrect."** As explained further below, these additional [redacted] logs also do not change this opinion because the maybe_chrome_incognito bit (and the is_chrome_non_incognito_mode bit) in these logs is still based on the absence of the X-Client-Data header. As I explained in my Report, this is not a reliable method for accurately identifying the use of Incognito mode because there are instances where the X-Client-Data header is not sent, but the user is not browsing in Incognito mode. These "false positives" render the maybe_chrome_incognito and is_chrome_non_incognito_mode bits unreliable for the reasons stated in my Report, and the existence of additional logs that contain the same bits does not change this opinion. *Id.* ¶¶ 142-145.

29. **Report Section III.J (Opinion 10): "Mr. Hochman's Proposed Methods For Identifying Class Members (# 22) Do Not–And Cannot–Account For Shared Devices Or Accounts."** The additional [redacted] logs do not change this opinion because the same issues described in my Report will also affect any attempt to identify class members by applying the same fingerprinting methodology to these logs. *Id.* Specifically, Plaintiffs' proposed methodology still does not account for shared devices, which render their methodology unreliable in light of widespread device-sharing and resulting collisions of IP address and user agent combinations. *See Id.* ¶¶ 162-180.

30. As noted above, the [redacted] additional logs containing the maybe_chrome_incognito bit identified in the June 14 Šrámek Declaration also do not change the opinions stated in my June 7,

2022 Expert Report submitted in this case regarding the "Incognito detection bits" unsuitability for identifying individual users of Incognito mode. *See, e.g.*, Dkt. 659-10 § III.H. The existence of additional logs containing the maybe_chrome_incognito bit does not change my opinion that Plaintiffs' Expert Jonathan "Hochman's opinion that the 'maybe_chrome_incognito' bit reliably detects Incognito traffic is incorrect" because that opinion is based on the fundamental way that the maybe_chrome_incognito bit is computed. *Id.* As I stated in my Report, the maybe_chrome_incognito bit is a "boolean field that relies on the absence of the X-Client-Data header to approximate and monitor traffic Google receives from Chrome instances in Incognito mode," and "the absence of the X-Client-Data header cannot be used to reliably detect Incognito traffic because there are a variety of cases in which the X-Client-Data header *is not* sent by a Chrome browser when a user is using a browser in non-Incognito mode (false positives)." *Id.* ¶ 143. As such, the maybe_chrome_incognito bit in these additional [redacted] logs also "cannot be used to reliably detect Incognito traffic because there are a variety of cases in which the X-Client-Data header *is not* sent by a Chrome browser when a user is using a browser in non-Incognito mode (false positives)." *Id.* (emphasis in original). In other words, these additional [redacted] logs also "can not be used to reliably detect Incognito traffic, let alone identify purported members of Class I" because they are affected by the same false positive problem as the previously-identified logs. *Id.* ¶ 145. And for these additional [redacted] logs, there is also "no way to exclude the false positives . . . because the reason for false positives is 'not observable from a server perspective.'" *Id.* (quoting Berntson June 16, 2021 Tr. 384:23-24).

31. Similarly, the [redacted] log identified in the June 14 Šrámek Declaration does not change my opinions regarding the is_chrome_non_incognito_mode bit stated in my Report. *See, e.g.*, Dkt. 659-10 § III.H. As I explained in my Report, the is_chrome_non_incognito_mode bit also relies on the absence of the X-Client-Data header to

approximate Incognito traffic. *Id.* ¶ 143 & n. 172. As such, it suffers from the same flaws described in paragraphs 28 and 30 above.

32. **Report Section IV.C (Opinion 13): "Mr. Schneier's Assertion That Google Has Not Taken Steps To Ensure That A User's Choice To Sign Out Of A Google Account Will Prevent Google From Associating The User's Signed-Out Activity With Any Signed-In Data Is Incorrect."** The [redacted] log described above also does not change my opinions regarding the steps Google has taken to prevent joining of authenticated and unauthenticated information. As described above, [redacted] records (*i.e.*, adds records from input logs to an output log without joining any of the records contained in either log). The code described above provides an additional example of Google's technical restrictions designed to prevent the joining of authenticated and unauthenticated information, and thus it reinforces and further supports this opinion.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on the 30th day of November 2022 at Irvine, CA.

By: DocuSigned by: Konstantinos Psounis EDAADC53351F49E...
Konstantinos Psounis, PhD